Ann. Hum. Genet. (2000), **64**, 443–453
Printed in Great Britain

On power and efficiency robust linkage tests for affected sibs

J. L. GASTWIRTH[1] AND B. FREIDLIN[2]

[1] *Division of Cancer Genetics and Epidemiology, EPS 8105, National Cancer Institute, Bethesda, MD 20892, USA*

[2] *Division of Cancer Treatment and Diagnosis, National Cancer Institute, USA*

(Received 10.1.00. Accepted 5.6.00)

SUMMARY

For diseases that do not follow a clear Mendelian pattern of inheritance nonparametric tests applied to affected sibs have been shown to be robust to the inherent uncertainty about the precise underlying genetic model. It is known that the weights optimizing the power of tests using IBD alleles shared by affected sib pairs or triples depend on the underlying model. We show how efficiency robustness techniques, used in other areas of statistics, provide a systematic approach for constructing a robust linear combination of the statistics that are optimal for the individual members of a family of plausible genetic models. The method depends on the correlation matrix of the optimal tests as these correlations reflect how different the models are. When the minimal correlation is less than 0.5, an alternate robust procedure is proposed. The methods apply to combining data from sibships of different sizes.

INTRODUCTION

Classical likelihood linkage analysis assumes a parametric model relating the effects of genotypes at a single locus to a trait phenotype (Ott, 1999). This method has contributed to the identification of highly penetrant genes for diseases following simple Mendelian patterns, e.g. Huntington disease and Alzheimer's disease. For more complex diseases, for which the precise model of inheritance may not be known, nonparametric methods that require fewer model assumptions have been developed (Blackwelder & Elston, 1985; Risch, 1990; Faraway, 1993; Holmans, 1993). These have been shown to perform reasonably well under a broad range of genetic models (Kruglyak *et al.* 1996; Teng & Siegmund, 1997; Olson *et al.* 1999). Whittemore & Tu (1998) proposed a robust 'minimax' test that has good power properties over a family of models. We show that the theory of efficiency robustness (Gastwirth, 1966; Birnbaum & Laska, 1967; Gastwirth 1985; Lachin & Wei, 1988; Burnett *et al.* 1989; Zucker & Lakatos, 1990; Podgor *et al.* 1996; Broet *et al.* 1999; Freidlin *et al.* 1999) yielding the maximin efficiency robust test (MERT) is applicable to a wide class of genetic models including those considered by Schaid & Nick (1990), Sham, Zhao & Curtis (1997) and Whittemore & Tu (1998). The approach is based on the correlation matrix of the optimal tests, as the squares of the correlations are their relative asymptotic efficiencies at each of the possible models. For affected sib-pairs the MERT test is the robust linear combination of the means and proportions tests obtained by Whittemore & Tu (1998). A non-linear robust test based on the maximum of the optimal tests is also discussed and conditions when it is preferable to the MERT are given. The maximum of the means and proportions tests for affected sib-pairs was considered by Schaid & Nick (1990) but it is not noticeably more powerful than the maximin efficient linear combination (MERT) of the two

Correspondence: J. L. Gastwirth, Department of Statistics, George Washington University, Washington, DC 20052, USA.
E-mail: jlgast@gwu.edu

statistics. For broader families of models, however, the maximum of several optimum procedures can be more powerful than the MERT. Indeed, this is the case for affected sib-triples. Throughout the paper we restrict attention to a single marker for which the IBD status of siblings can be clearly determined. The methods of constructing the MERT should readily extend to the situation where the IBD status is estimated as only the correlations of the optimal tests will change.

GENERAL BACKGROUND

In order to compare the powers of two consistent tests of the null hypothesis $\theta = \theta_0$ vs. $\theta > \theta_0$, Pitman (Noether, 1955; Stuart & Ord, 1991) introduced the concepts of asymptotic efficiency and efficacy. Let two tests T_1 and T_2 have a large sample bivariate normal distribution. One evaluates the limiting power of the tests for a sequence of alternatives $\theta_n = \theta_0 + k/(n^{1/2})$, that approach the null hypothesis as the sample size, n, increases (k is a constant). The asymptotic power of the tests T_i $(i = 1, 2)$ against the alternative θ_n can be shown to be

$$\Phi(c_i k - z_\alpha), \tag{1}$$

where z_α is the $100(1-\alpha)$ percentile of the standard normal distribution and the efficacy

$$c_i = \lim_{n_i \to \infty} \frac{E'_{\theta_0}(T_i)}{\sqrt{V_{\theta_0}(T_i)} n_i^{1/2}}. \tag{2}$$

$E_\theta(T_i)$ and $V_\theta(T_i)$ are mean and variance of T_i, which are functions of θ and Φ is the c.d.f. of a standard normal distribution. The asymptotic relative efficiency (ARE) of test T_1 to test T_2 is the ratio of the sample sizes required for the tests to have equal asymptotic power. From (1) and (2):

$$\mathrm{ARE}(T_1, T_2) = \lim \left(\frac{n_2}{n_1} \right) = \left(\frac{c_1}{c_2} \right)^2. \tag{3}$$

Suppose one of the tests, say T_1, is the asymptotically most powerful for a particular model. The ARE of another consistent test T_2 to T_1 is the square of their correlation under the null hypothesis. In many applications including linkage analysis the data may be assumed to arise from one of several scientifically plausible genetic models. Often the optimal tests T_i for each model are asymptotically jointly normally distributed with correlation matrix $\{\rho_{ij}\}$. Then the Pitman ARE of the test T_j relative to the test T_i when T_i is optimum is $\rho_{ij}^2 = \langle T_i, T_j \rangle^2$. Thus, from (3), the sample size needed by the sub-optimal test T_j relative to that needed by T_i to achieve the same power that T_i has is given by

$$\lim \left(\frac{n_j}{n_i} \right) = 1/\rho_{ij}^2. \tag{4}$$

It should be noted that for 1 D.F. tests such as those developed by Whittemore & Halpern (1994), Teng & Siegmund (1997) and Whittemore & Tu (1998),

$$\rho^2 = \frac{\zeta'^2}{\zeta^2}.$$

where ζ (ζ') is the non-centrality parameter of the optimal (other) test. Whittemore & Tu's penalty, which expresses the percentage increase in sample size needed by the non-optimal test to achieve the same power as the optimal one, is $\rho^{-2} - 1$.

Suppose one has I possible models with corresponding optimal tests Z_i $(i = 1, \ldots, I)$, standardized versions of T_i (mean 0 and variance 1). For a test Z denote its relative efficiency under the model i

by $e(Z,i) = \langle Z, Z_i \rangle^2 = \rho_i^2$. The minimum value, $e^*(Z)$, of $e(Z,i)$ over the models measures the greatest loss in power of test Z relative to the optimal tests. The maximin efficiency robust test (MERT), M, satisfies $e^*(M) = \sup_{Z \in \Gamma} e^*(Z)$ where Γ is the set of all consistent asymptotically normal tests for the problem, i.e. M has the highest minimum efficiency. Gastwirth (1966) showed that when the minimum correlation, $\rho^* = \min(\rho_{ij})$, of the optimal tests Z_i, is > 0 the MERT exists, is unique, and is a linear combination of the $\{Z_i\}$. Note that the test yielding the MERT, maximizing the minimum value of ρ_{ij} for a set of models, is also the test minimizing the maximum penalty (loss of relative efficiency) described by Whittemore & Tu.

A simple algorithm to obtain the MERT is given in Gastwirth (1985) and is used by Zucker & Lakatos (1990) to analyse survival data. The idea is to find the most disparate models, i.e. those whose corresponding tests Z_1 and Z_2 have correlation $\rho_{12} = \rho^*$ (the minimum). The MERT for these two models is given by $R_2 = (Z_1 + Z_2)/\sqrt{2(1+\rho_{12})}$ and has efficiency $(1+\rho_{12})/2$ for each of them. To see whether R_2 is the MERT for the entire family one computes its correlation with the other optimal Z_i. If all of them are at least ρ^*, R_2 is the MERT for the entire family. Otherwise, there will be a test, Z_3, having the lowest correlation with R_2. The linear combination, R_3, of Z_1, Z_2 and Z_3, which has equal correlation (ρ^{**}) with each of them is their MERT. If the correlation of R_3 with each remaining Z_i is at least ρ^{**}, R_3 is the MERT for the entire family. If not, one repeats the process.

Although the assumption of a finite set of models may appear to limit the applicability of the MERT approach, Gastwirth & Podgor (1992) demonstrated that the MERT statistic, M, for $\{Z_i\}$ remains the MERT for the family of all convex combinations of $\{Z_i\}$. A routine calculation shows that if one uses non-standardized versions of the optimal tests, any convex combination of them is expressible as a multiple of another convex combination of the standardized optimal tests. Hence, the MERT, M, of the $\{Z_i\}$ or $\{T_i\}$ is the MERT for all tests in the closed convex hull of either defining family.

Another robust test statistic is $\text{MAXI} = \max_{1 \leqslant i \leqslant I} (Z_i)$. Asymptotically, under the null hypothesis, MAXI is distributed as $\max [MN(0, \{\rho_{ij}\})]$. For two plausible models for survival data this statistic was proposed by Tarone (1981) and a similar procedure for analysis of ASP data is discussed by Schaid & Nick (1990). In the context of survival analysis, the maximum of several tests has been used by Fleming & Harrington (1991) and Lee (1996). Freidlin, Podgor & Gastwirth (1999) considered survival and categorical data models and showed that when $\rho^* \leqslant 0.5$ the MAXI test is more powerful than the MERT but when $\rho^* \geqslant 0.7$ there was virtually no difference in their powers.

The correlation matrix, $\{\rho_{ij}\}$, of the optimal statistics summarizes the structure of the family of alternative models as each correlation reflects how close, statistically, the two models are. More precisely, the distance between the standardized optimal tests, $\|T_i - T_j\|$, equals $2(1-\rho_{ij})$. The advantage of the efficiency robustness approach is that the optimal tests for a wide variety of statistical problems have a joint asymptotically normal distribution so robust tests can be obtained routinely.

Efficiency robust tests have high efficiency relative to the optimal tests for each of the possible models. They should be distinguished from the tests that possess the largest minimum power for testing linkage over the same class of models. This criterion was introduced in the nonparametric setting by Doksum (1966, 1967) who standardized the various alternative models for two sample rank tests, e.g. by requiring the distributions to have the same variance. Gastwirth & Podgor (1992) noted that if one standardizes the models by their Fisher information the two criteria then agree. The test optimal for the least informative model often satisfies the maximin power criteria but not the maximin efficiency one. The difference between the two criteria will be illustrated in the discussion of sib-pairs.

AFFECTED SIB-PAIRS

Following Schaid & Nick (1990) let $N = (n_0, n_1, n_2)$ denote the numbers of sib-pairs that inherit 0,1 and 2 marker alleles IBD, respectively ($n = n_0+n_1+n_2$). Denote the vector of probabilities that a sib-pair shares 0, 1 and 2 marker genes by $P = (p_0, p_1, p_2)$. The empirical estimate of P is

$$\hat{P} = (\hat{p}_0, \hat{p}_1, \hat{p}_2) = \left(\frac{n_0}{n}, \frac{n_1}{n}, \frac{n_2}{n}\right).$$

Since the random vector N has a trinomial distribution with parameters (n, p_0, p_1, p_2), it can be shown (Rao, 1973) that $\sqrt{n}(P-\hat{P}) \sim N(\mathbf{0}, \text{diag}\,(P)-P'P)$. Under the null hypothesis of no linkage $P = (\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$. Various plausible genetic models lead to different optimal tests Z_i because they yield different IBD proportions (Suarez *et al.* 1978). The finite sample and asymptotically optimal tests against a specific alternative are given in Knapp (1991) and Schaid & Nick (1990). Let W_i denote the optimal weight vector for the ith model. Then the optimal test T_i and its standardized versions Z_i are

$$T_i = W_i'(\hat{P}-P) \quad \text{and} \quad Z_i = \frac{\sqrt{n}W_i'(\hat{P}-P)}{\sqrt{W_i'(\text{diag}\,(P)-P'P)W_i}} \tag{5}$$

The null correlation between the tests optimal for models i and j is

$$\rho_{ij} = \frac{W_i'(\text{diag}\,(P)-P'P)W_j}{\sqrt{W_i'(\text{diag}\,(P)-P'P)W_i}\sqrt{W_j'(\text{diag}\,(P)-P'P)W_j}}.$$

Whittemore & Tu (1998) considered the family $F = \{F_a: 0 \leqslant a \leqslant 0.5\}$ of models for the underlying IBD sharing probabilities where

$$F_a = \left[P:\left(\lambda(0, a, 1-a)+(1-\lambda)\left(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}\right)\right)\right]. \quad (0 \leqslant \lambda \leqslant 1) \tag{6}$$

They showed that the efficient score test, T_a, for a model F_a is of the form (5) with

$$W = \left(0, \frac{a/2}{1-a}, 1\right). \tag{7}$$

The well known means test, $T_{0.5}$, is based on $0.5\hat{p}_1+1\hat{p}_2$ ($W = \{0, 0.5, 1\}$), and the proportions test T_0 based on $\hat{p}_2$ ($W = \{0, 0, 1\}$) are optimal for the models a = 0.5 and a = 0, respectively (see Whittemore & Tu, 1998). The null correlation of the tests T_0 and $T_{0.5}$ is 0.8165 and corresponding MERT is $(Z_0+Z_{0.5})/\sqrt{2(1.8165)}$ where $Z_{0.5} = (0.5\hat{p}_1+\hat{p}_2-1/2)/\sqrt{2/(16n)}$ and $Z_0 = (\hat{p}_2-1/4)/\sqrt{3/(16n)}$. In terms of (7) the MERT is specified by $W = \{0, 0.275, 1\}$. Note that for any model F_a in F, its optimal test, T_a is the convex combination $\frac{a}{1-a}T_{0.5}+\left(1-\frac{a}{1-a}\right)T_0$. Thus, the MERT for the pair $(T_0, T_{0.5})$ is the MERT for the entire family F and has the same maximin efficiency, $(1+\rho)/2$, relative to the optimal tests for any of the models.

Another possible procedure is the maximum of T_0 and $T_{0.5}$, which we will call MAX2, and was discussed by Schaid & Nick (1990). Sham (1998, p.116) suggested using the more significant of T_0 and $T_{0.5}$ and doubling the p-value.

We conducted a power simulation to compare the properties of the various optimal tests and the

Table 1. *Empirical power estimates for the sib-pair tests (for sample size 200)*

Simulate under	Proportions test $W =$ $(0, 0, 1)$	Optimal test $W =$ $(0, 0.167, 1)$	Optional test $W =$ $(0, 0.333, 1)$	Means test $W =$ $0, 0.5, 1)$	MERT $W =$ $(0, 0.275, 1)$	Max2 (Ex. Pair)	2*Min p-value
Null	0.0406	0.0493	0.0474	0.0509	0.0491	0.0460	0.0406
$W = (0, 0, 1)$	0.805	0.813	0.769	0.681	0.793	0.785	0.779
$W = (0, 0.167, 1)$	0.771	0.806	0.790	0.745	0.802	0.779	0.767
$W = (0, 0.333, 1)$	0.711	0.783	0.800	0.797	0.799	0.780	0.761
$W = (0, 0.5, 1)$	0.577	0.698	0.757	0.798	0.740	0.752	0.725

The NULL row gives the estimated size of a nominal 0.05 level test. Because they differ slightly from 0.05, the power of the optimal test may sometimes be lower than a test with which it is highly correlated.
Estimates are based on 1000000 replications.

Table 2. *Empirical power estimates of means, proportions and MERT tests when $\lambda = 0.0518$*

	$(p_0, p_1, p_2) = (0.237, 0.474, 0.289)$	$(p_0, p_1, p_2) = (0.237, 0.5, 0.263)$
Proportions	0.804	0.168
Means	0.642	0.217
MERT	0.753	0.202

Estimates are based on 1000000 replications of samples of 1000. The large sample size was used to ensure that the asymptotic approximation to the 0.05 level was very accurate.

three robust tests. Table 1 presents the powers, obtained by simulation, of the robust tests and four optimal tests specified in (5) and (7) for different members of the family $\{T_a\}$, $0 \leqslant a \leqslant 0.5$. The 0.05 level of significance was used. The parameter λ in (6) was selected so that the power of the optimal test was approximately 80%. The results in Table 1 indicate that there can be a substantial loss in power when the means test is used in situations where the proportions test is optimal and vice-versa. The MERT and MAX2 efficiency robust procedures provide similar protection against loss of power relative to the optimal test when the model is uncertain. Sham's method can be considered as a useful approximation to the MAX2 test as it is slightly less powerful. Both robust procedures have their largest relative loss of power at the two extreme genetic models. The sample size required by the MERT can be obtained from relationship (4). If the optimal test is used to analyse 100 sib-pairs this implies that a sample of 110 is needed for the MERT when the data come from one of the extreme models, in agreement with Table 1 in Whittemore & Tu (1998).

To appreciate the difference between the two criteria, maximin power and maximin efficiency, consider the properties of the means and proportions tests for the models $F_{0.5}$ and F_0 when $\lambda = 0.0518$. The corresponding alternative trinomial distributions for which the means and proportions tests are optimal are $(0.237, 0.5, 0.263)$ and $(0.237, 0.474, 0.289)$. The first alternative, corresponding to the model $F_{0.5}$, is closer to the null value $(0.25, 0.5, 0.25)$. The powers of the two tests are given in Table 2. The means test has maximin power, however, one is trading a loss of 15% of power against the first model for a 4% gain against the second model. Note that the MERT has high power relative to the optimal tests for both models.

AFFECTED SIB-TRIPLES

In this section we apply efficiency robustness principles to tests relying on n affected sib-triples. Following the development given in Whittemore & Tu (1998), there are four possible IBD configurations for three siblings (Sribney & Swift, 1992; Feingold *et al.* 1993). They are denoted by $z_i, i = 0, \ldots, 3$ and their sample proportions by $\hat{z}_i$. The 1 D.F. tests for sib triples is based on weighted

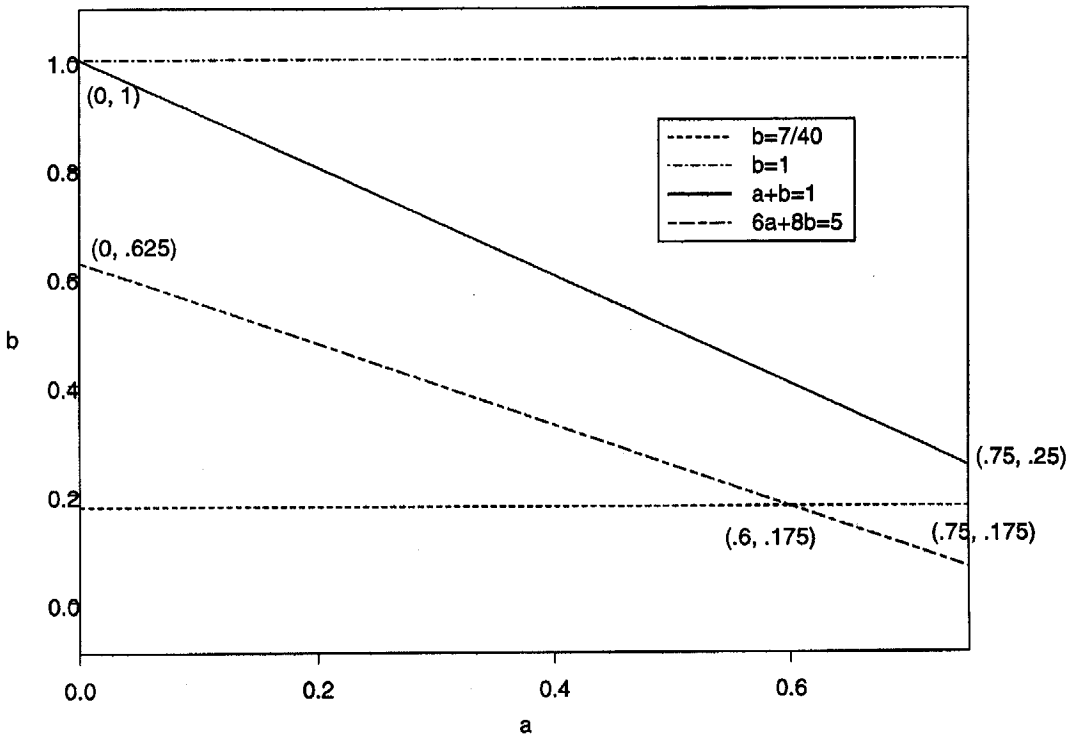

Figure 1.

sums of the $\hat{z}_i = W'\hat{Z}$ where $W = (w_0, w_1, w_2, w_3)$ and $\hat{Z} = (\hat{z}_0, \hat{z}_1, \hat{z}_2, \hat{z}_3)$. As the $\hat{z}_i$ sum to 1 the standardized statistics can be expressed as

$$Z = \frac{\sqrt{n}\left[w_0\left(\hat{z}_0 - \frac{3}{8}\right) + \hat{z}_1 - \frac{1}{16} + w_2\left(\hat{z}_2 - \frac{3}{8}\right)\right]}{\sqrt{\frac{15}{64}(w_0^2 + w_2^2) - \frac{9}{32}w_0 w_2 - \frac{6}{128}(w_0 + w_2) + \frac{15}{256}}} \tag{8}$$

Whitemore & Tu (1998) consider the family of models

$$F_{ab} = \left[z = (z_0, z_1, z_2, z_3) = \lambda(a, b, 0, 1-a-b) + (1-\lambda)\left(\frac{3}{8}, \frac{1}{16}, \frac{3}{8}, \frac{3}{16}\right)\right],$$

for $0 \leqslant \lambda \leqslant 1$ and (a, b) satisfying the following constraints

$$\left.\begin{aligned} 0 \leqslant a \leqslant \tfrac{3}{4} \\ \tfrac{7}{40} \leqslant b \leqslant 1 \\ a + b \leqslant 1 \\ 6a + 8b \geqslant \end{aligned}\right\} \tag{13}$$

Whittemore & Tu showed that the coefficients (w_0 and w_2) of the optimal test (8) for the model F_{ab} are given by

$$w_0 = \frac{3a/2 + b - 1}{a + 4b - 1}, \quad w_2 = \frac{a + b - 1}{a + 4b - 1} \tag{10}$$

The genetically extreme models lie at the edges of the region (9) in the (a,b) space which is given in Figure 1. The five extreme models are vertices of this region. As every interior point of the region

Table 3. *Empirical power estimates for the sib-triples tests (sample size 200)*

Simulate under	$W =$ (0.5, 1, 0, 0)	$W =$ (0, 1, 0, 0)	$W =$ (0.67, 1, −0.17, 0)	$W =$ (−0.25, 1, −0.25, 0)	$W =$ (0.25, 1, −0.75, 0)	MERT	Max5
Null	0.0563	0.0387	0.0514	0.0519	0.0511	0.0493	0.0469
$W = (0.5, 1, 0, 0)$	0.798	0.395	0.771	0.317	0.675	0.674	0.713
$W = (0, 1, 0, 0)$	0.534	0.801	0.324	0.796	0.424	0.724	0.761
$W = (0.67, 1, -0.17, 0)$	0.778	0.247	0.797	0.208	0.722	0.616	0.712
$W = (-0.25, 1, -0.25, 0)$	0.362	0.734	0.211	0.805	0.426	0.667	0.722
$W = (0.25, 1, -0.75, 0)$	0.683	0.319	0.720	0.379	0.800	0.686	0.713

The Null row gives the estimated size of a nominal 0.05 level test.
Estimates are based on 1000000 replications.

Table 4. *Sample size required for same power as the optimal test has for 100 observations*

	Tests			
Optimal weights	$W = (0.5, 1, 0, 0)$	$W = (0, 1, 0, 0)$	MERT	Max5
$W = (0.5, 1, 0, 0)$	100	250	126	120
$W = (0, 1, 0, 0)$	250	100	137	119
$W = (0.67, 1, -0.17, 0)$	106	472	146	117
$W = (-0.25, 1, -0.25, 0)$	400	111	147	121
$W = (0.25, 1, -0.75, 0)$	133	333	123	120

is a convex combination of the vertices, the family of models F_{ab} is the convex hull of the five extreme models. Thus the MERT for these five models is also the MERT for the entire family. The correlation matrix of the optimal tests with the corresponding values of (w_0, w_2) is:

(0.5, 0)	(0, 0)	(0.67 −0.17)	(−0.25, −0.25)	(0.25, −0.75)
1.000	0.632	0.970	0.500	0.866
0.632	1.000	0.460	0.949	0.548
0.970	0.460	1.000	0.364	0.910
0.500	0.949	0.364	1.000	0.577
0.866	0.548	0.910	0.577	1.000

As $\rho_{34} = 0.364$ is the minimum correlation, for this family, the third and fourth models are the 'extreme pair'. The MERT coefficients are: (0, 0, 0.6055, 0.6055, 0), corresponding to $w_0 = 0.138$ and $w_2 = -0.212$. The MERT has maximum efficiency 0.682 and AREs 0.792, 0.728, 0.682 0.682 and 0.811 relative to the optimal tests for the five models. An alternative robust test is MAX5, the maximum of the standardized optimal tests for the five extreme models. Table 3 presents results of a power simulation study for the sib-triple tests. The empirical power estimates are tabulated for the optimal tests for the five models and the two robust tests. The data for samples of size 200 were simulated under the five extreme models.

The results in Table 3 show that there can be a substantial loss of power when the test optimal for a genetical model different from the true one is used. The greatest such loss (0.797−0.208 = 0.589 or 0.805−0.211 = 0.594) occurs when the test optimal for the third (fourth) model is used when the true underlying model is fourth (third). As the minimum null correlation between the optimal test statistics was 0.364, the results of Freidlin *et al.* (1999) suggest that the MAX5 test will perform better than the simpler MERT. The last two columns confirm this, the additional power of MAX5 ranged from 0.03 to 0.09. Both methods, however, provide reasonable power protection.

Another way of presenting the advantage of the robust procedures is to look at the sample sizes needed to have a desired power under various alternatives. Table 4 gives the sample sizes required

to reach the same asymptotic power as the optimal test. The table is based on the asymptotic results for the two optimal tests and the MERT. The values for the MAX5 tests were obtained empirically. The lower sample size needed by MAX5 relative to the MERT reflects its increased power robustness in this situation. Note that the sample sizes for the optimal tests and MERT agree with Table 4 of Whittemore & Tu (1998).

COMBINING DATA FROM DIFFERENT TYPES OF SIBSHIPS

Sham *et al.* (1997) considered n sibships with $a_1, \ldots, a_n$ affected members and $u_1, \ldots, u_n$ unaffected members and developed a method for combining the results of sibships of the various (a_i, u_i) types. When the sample contains a number of sibships with three or more affected members another complication arises: an affected sibship contributes $a(a-1)/2$ sib-pairs pair but only $a-1$ of these $a(a-1)/2$ pairs are independent. Thus sib-pairs from sibships of different size and different number of affected sibs should be weighted appropriately to reflect the relative amount of information they contribute. Several authors (Suarez & Hodge, 1979; Hodge, 1984) proposed weighting schemes based on various measures of information. Sham *et al.* (1997) developed an approach that yields the test with the maximum asymptotic power for a particular single locus model. They also considered the situation where the data followed one of five models: rare recessive (RR), rare dominant (RD), common recessive (CR), common dominant (CD) and Alzheimer's (AZ) that represents a minor susceptibility locus and used the means tests for each sibship (l = (0, 1, 2)). A robust combination of these optimal tests will be obtained using the results in Sham *et al.* (1997).

Denote the number of sibships in the sample with a affected and u unaffected members by k_{au}. Sham *et al.* obtained the optimum weights, w_{au}, for each type of sibship under the five different genetic models, their combined statistic is

$$T = \frac{\sum_{(a,u)} w_{au} \sum_{i=1}^{k_{au}} (Y_i - m_i)}{\sqrt{\sum_{(a,u)} k_{au} w_{au}^2 s_a^2}},$$

where Y_i is the sum of number of alleles IBD in the sibship i; m_i and s_i^2 are its mean and variance under the null hypothesis.

The null correlation between two optimal tests with weights w_{au} and v_{au}, respectively, is:

$$\rho_{wv} = \frac{\sum_{(a,u)} w_{au} v_{au} k_{au} s_a^2}{\sqrt{k_{au} w_{au}^2 s_a^2} \sqrt{k_{au} v_{au}^2 s_a^2}}. \tag{11}$$

To illustrate the efficiency robust methodology we will assume the following data, where (a,u) denotes the number of affected and unaffected sibs in a sibship: twenty sibships of each of the three types (2, 0), (2, 1) and (2, 2) and 10 sibships of type (7, 0). Using formula (11) and weights in Table 6 of Sham *et al.* (1997) the correlation matrix of the optimal tests is

1	0.755	0.998	0.879	0.908
0.755	1	0.713	0.977	0.959
0.998	0.713	1	0.847	0.880
0.879	0.977	0.847	1	0.997
0.908	0.959	0.880	0.997	1

The second and third models (rare dominant (RD) and common recessive (CR)) were the extreme pairs. As the minimum correlation was 0.713 the MERT is an appropriate robust procedure. It gives

each member of the extreme pair a weight of 0.54 and its maximin efficiency is 0.857. This implies that if one used the optimal test for the RD model when the true model was CD, then one would need $100/(0.713)^2 = 197$ families, assuming a similar proportion of sibships of the various types. The corresponding sample size using the MERT is 117.

Feingold *et al.* (1993) showed that triples should receive weight $c:1$ relative to pairs when data from triples and pairs are available. The optimal value of c depends on the underlying genetic model. The efficiency robust approach uses the optimal tests utilizing both triples and pairs for each genetic model. For the ith model, let P_i and T_i denote the optimal tests for pairs and triples, respectively and let c_i be the weight yielding the optimal test $S_i = P_i + c_i T_i$ with standardized form $z_i = (P_i + c_i T_i)/\sqrt{V(P_i) + c_i^2 V(T_i)}$, where $V(P_i)$ and $V(T_i)$ are variances of the statistics P_i and T_i, respectively. The correlations between the Z_i can now be computed as in (11) and the MERT and MAX obtained. As before the minimum correlation will assist in deciding between the MERT and the MAX. Our approach differs from that of Whittemore & Tu (1998) who combined robust tests for triples and pairs by choosing a compromise value of c.

DISCUSSION

The problem of selecting a set of weights for a test based on IBD sharing so that it has high efficiency for a range of scientifically plausible models of inheritance has been discussed by Kong & Cox (1997), Sham *et al.* (1997) and Whittemore & Tu (1998). The choice of models should rely on previous research. For instance, Schaid & Sommer (1994) indicated that molecular biology can assist in the process. We have demonstrated that an approach used to obtain robust tests for a variety of statistical problems applies to linkage tests. Recall that the non-parametric feature of tests based on IBD sharing is similar to that of nonparametric rank tests. In both settings the distribution of a wide class of test statistics, *under the null hypothesis*, is the same. The optimal test against a particular alternative model, however, depends on that model. Thus one needs to obtain a compromise method that has high relative efficiency across a set of models. The value ρ^* reflects how disparate the plausible underlying models are as the largest minimum efficiency possible for a test using linear combinations of the optimal tests is $(1+\rho^*)/2$.

The efficiency robust methods described here can be used when the IBD configuration of some sibs cannot be unambiguously determined or when the marker locus does not coincide with the disease locus. All that is required is that the optimal tests for each model have a joint large-sample normal distribution as the matrix of squared correlations gives the corresponding AREs.

Because the same uncertainty about the underlying genetic model arises when many markers are tested for linkage (Kruglyak *et al.* 1996; Feingold & Siegmund, 1997; Teng & Siegmund, 1997), the methodology of this paper can be used to select a suitable robust test that can be applied at all markers. Then the appropriate distribution of the supremum of these correlated tests can be used. As the MERT is a linear combination of the allele sharing counts of each type of pedigree, it has an asymptotic normal distribution so the methods of Teng & Siegmund (1997) are directly applicable. If the correlation matrix of the optimum tests for the plausible underlying genetic models indicates that the MAX of the tests optimal for the extreme models should be used, then the distribution of the robust procedure will need to be simulated as the supremum of correlated sets of maxima of even two normal distributions is not known. These considerations also apply to the affected pedigree (APM) tests (Weeks & Lange,1988; Whittemore, 1996) which are weighted sums of pedigree counts with various IBD configurations. Again the optimal weights depend on the underlying genetic model.

We thank Lynn Goldin for several useful discussions and the referees for helpful suggestions. This research was partially supported by a grant from the National Science Foundation.

REFERENCES

Birnbaum, A. & Laska, E. (1967). Efficiency robust two-sample rank tests. *Journal of the American Statistical Association* **62**, 1241–1257.

Blackwelder, C. & Elston, R. C. (1985). A comparison of sib-pair linkage tests for disease susceptibility loci. *Genet. Epidemiol.* **2**, 85–97.

Broet, P., Moreau, T., Lellouch, J., Asselain, B. (1999). Unobserved covariates in the two-sample comparison of survival times: a maximin efficiency robust test. *Statistics in Medicine* **18**, 1791–1800.

Burnett, R., Krewski, D., Bleuer, S. (1989). Efficiency robust score test for rodent tumourigenicity experiments. *Biometrika* **76**, 17–324.

Doksum, K. (1966). Asymptotically minimax distribution-free procedures. *The Annals of Mathematical Statistics* **37**, 619–628.

Doksum, K. (1967). Asymptotically optimal statistics in some models with increasing failure rate average. *The Annals of Mathematical Statistics* **38**, 1731–1739.

Faraway, J. J. (1993). Improved sib-pair linkage test for disease susceptibility loci. *Genet. Epidemiol.* **10**, 225–233.

Feingold, E., Brown. P. O., Siegmund, D. (1993). Gaussian models for genetic linkage analysis using complete high resolution maps of identity by descent. *Am. J. Hum. Genet.* **53**, 234–251.

Feingold, E. & Siegmund, D. O. (1997). Strategies for mapping heterozygous recessive traits by allele-sharing methods. *Am. J. Hum. Genet.* **60**, 965–978.

Fleming, T. R. & Harrington, D. P. (1991). *Counting processes and survival analysis*. Wiley, New York.

Freidlin, B., Podgor, M. J., Gastwirth, J. L. (1999). Efficiency robust tests for survival or ordered categorical data. *Biometrics* **55**, 883–886.

Gastwirth, J. L. (1966). On robust procedures. *Journal of the American Statistical Association* **61**, 929–948.

Gastwirth, J. L. (1985). The use of maximin efficiency robust tests in combining contingency tables and survival analysis. *Journal of the American Statistical Association* **80**, 380–384.

Gastwirth, J. L. & Podgor, M. J. (1992). Efficiency robust rank tests for the location scale problem. In *Nonparametric statistics and related topics*. Amsterdam: Elsevier (ed. A. E. Saleh), pp. 17–31.

Hodge, S. E. (1984). The information contained in multiple sibling pairs. *Genet. Epidemiol.* **1**, 109–122.

Holmans, P. (1993). Asymptotic properties of the affected-sib-pair linkage analysis. *Am. J. Hum. Genet.* **52**, 362–374.

Knapp, M. (1991). A powerful test of sib-pair linkage for disease susceptibility. *Genet. Epidemiol.* **8**, 141–143.

Kong, A. & Cox, N. J. (1997). Allele-sharing models: LOD scores and accurate linkage tests. *Am. J. Hum. Genet.* **61**, 1179–1188.

Kruglyak, L., Daly, M. R., Reeve-Daly, M. P., Lander, E. S. (1996). Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am. J. Hum. Genet.* **58**, 1347–1363.

Lachin, J. M. & Wei, L. J. (1988). Estimators and test in the analysis of multiple nonrespondent 2×2 tables with partially missing observations. *Biometrics* **44**, 513–528.

Lee, W. J. (1996). Some versatile tests based on the simultaneous use of weighted log-rank statistics. *Biometrics* **52**, 721–725.

Noether, G. E. (1955). On a theorem of Pitman. *Ann. Math. Statist.* **26**, 64–68.

Podgor, M. J. & Gastwirth, J. L. (1996). Efficiency robust rank tests for stratified data. In *Madan Puri Festschrift* (eds. E. Brunner & M. Denker), pp. 371–379. Utrecht: VSP.

Podgor, M. J., Gastwirth, J. L., Mehta, C. R. (1996). Efficiency robust tests of independence in contingency tables with ordered classifications. *Statistics in Medicine* **15**, 2095–2105.

Olson, J. M., Witte, J. S., Elston, R. C. (1999). Genetic mapping of complex traits. *Statistics in Medicine* **18**, 2961–2981.

Ott, J. (1999). *Analysis of human genetic linkage*, 3rd edn. Baltimore: Johns Hopkins University Press.

Rao, C. R. (1973). *Linear statistical inference and its applications*, 2nd edn. New York: Wiley.

Risch, N. (1990). Linkage strategies for genetically complex traits. II. The power of affected relative pairs. *Am. J. Hum. Genet.* **46**, 229–241.

Schaid, D. J. & Nick, T. G. (1990). Sib-pair linkage tests for disease susceptibility loci: common test vs. the asymptotically most powerful test. *Genet. Epidemiol.* **7**, 359–370.

Schaid, D. J. & Sommer, S. S. (1994). Comparison of statistics for candidate-gene association studies using cases and parents. *Am. J. Hum. Genet.* **55**, 402–409.

Sham, P. C., Zhao, J. H., Curtis, D. (1997). Optimal weighting for affected sib-pair analysis of sibship data. *Ann. Hum. Genet.* **61**, 61–69.

Sham, P. (1998). *Statistics in human genetics*. London: Arnold.

Sribney, W. M. & Swift, M. (1992). Power of sib-pair and sib-trio linkage analysis with assortative mating and multiple disease loci. *Am. J. Hum. Genet.* **51**, 773–784.

Stuart, A. & Ord, J. K. (1991). *Advanced theory of statistics*, vol. 2. New York: Oxford University Press.

Suarez, B. K., Rice, J., Reich, T. (1978). The generalized sib pair IBD distribution: its use in the detection of linkage. *Ann. Hum. Genet.* **42**, 87–94.

Suarez, B. K.& Hodge, S. E. (1979). A simple method to detect linkage for rare recessive disease: an application to juvenile diabetes. *Clinical Genetics* **15**, 126–136.

Tarone, R. E. (1981). On the distribution of the maximum of the log-rank statistics and the modified Wilcoxon statistics. *Biometrics* **37**, 79–85.

Teng, J. & Siegmund, D. (1997). Combining information within and between pedigrees for mapping complex traits. *Am. J. Hum. Genet.* **60**, 979–992.
Weeks, D. E. & Lange, K. (1988). The affected-pedigree-member method of linkage analysis. *Am. J. Hum. Genet.* **52**, 1200–1215.
Whittemore. A. S. & Halpern, J. (1994). A class of tests for linkage using affected pedigree members. *Biometrics* **50**, 118–127.
Whittemore, A. S. (1996). Genome scanning for linkage. *Am. J. Hum. Genet.* **59**, 704–716.
Whittemore, A. S. & Tu, I. P. (1998). Simple, robust linkage tests for affected sibs. *Am. J. Hum. Genet.* **62**, 1228–1242.
Zucker, D. & Lakatos, E. (1990). Weighted log-rank type statistics for comparing survival curves when there is a time lag in the effectiveness of treatment. *Biometrika* **77**, 853–864.